Econometric Methodology of Sales & Use Tax Base Forecasts

King County Office of Economic and Financial Analysis
July 13, 2010

The Office of Economic and Financial Analysis (OEFA) employs econometric models to generate sales & use tax base forecasts. The tax base includes all taxable retail sales and other "use" taxes in King County. Forecasts of the revenues going to King County are calculated from these tax base forecasts.

The forecasting model is of a two-step "error-correction" form, which has the practical benefit that it combines both cyclical and trend information in the same forecasting model. We first estimate a double-log form to get the trend equilibrium relationship and then embed it in a rate-of-change model to capture the cyclical turning points.

Step 1 – Estimation of the Trend Equilibrium Relationship

We will assume there is only one predictive variable X to keep the illustration simple.

$$lnY_t = \propto_1 lnX_{1t} + \propto_2 + u_t$$

Where:

$lnY_t =$	Natural log of sales & use tax base
$lnX_{1t} =$	Natural log of a predictive variable (e.g., personal income) which have available projections from a forecasting service.
$\propto_1, \propto_2 =$	Estimated coefficients
$u_t =$	Deviation of sales tax base from trend equilibrium at time *t*

We estimate this "cointegrating regression" by a technique called "fully modified least squares." [1] It captures the equilibrium relationship between a set of trending variables. The residual u_t measures the current deviation from equilibrium – if it is positive then the tax base is above its equilibrium level and will likely move downward until it reaches trend equilibrium. Likewise, if the residual is negative then revenue will likely move upward.

[1] Phillips and Hansen (1990), "Statistical Inference in Instrumental Variables Regression with I(1) Processes," *Review of Economic Studies,* 57, 99-125.

Step 2 – Estimation of the Error-Correction Model

The trend equilibrium relationship is important for out-year forecasts. However, that alone can miss cyclical turning points in the near-year forecasts. An "error-correction" model incorporates both trend and cycle information. We use rates of change to capture the cyclical turning points while using the residual estimated in Step 1 as the long run trend component:

$$\Delta lnY_t = \beta_1 \Delta lnX_{1t-1} + \beta_2 \Delta lnY_{t-1} + \beta_3 + \gamma u_{t-1} + \varepsilon_t$$

Where:

$\Delta lnY_t, \Delta lnY_{t-1}, \Delta lnX_{1t-1} =$	Variables from Step 1 in rate-of-change form
$u_{t-1} = lnY_{t-1} - \alpha_1 lnX_{1t-1} - \alpha_2 =$	Deviation from trend equilibrium last year
$\beta_1, \beta_2, \beta_3, \gamma =$	Estimated coefficients.

The coefficient γ governs the speed of adjustment back to equilibrium; it is expected to be between -1 and 0. If lnY_t is above its equilibrium value, then u is positive, and the negative γ will make the rate of change negative; so it will pull down or <u>correct</u> lnY_t back toward trend equilibrium. Likewise, if lnY_t is below its equilibrium value, then u is negative, and the negative γ will make the rate of change positive; so it will bump up or <u>correct</u> lnY_t back toward trend equilibrium.

Forecasting with the Error-Correction Model

Dynamic or "chain" forecasting log-levels from the error-correction model is straightforward. Let *T* be the end of the historical data, then the forecast *k* periods forward is:

$$lnY_{T+k} = lnY_{T+k-1} + \beta_1 \Delta lnX_{1T+k-1} + \beta_2 \Delta lnY_{T+k-1} + \beta_3 + \gamma[\, lnY_{T+k-1} - \propto_1 lnX_{1T+k-1} - \propto_2]$$

As long as we have forecasts for the predictive variable X, then Y can be dynamically updated into a chain of forecasts.

Variable Selection

The forecasting framework just illustrated depends on having forecasts of the predictive variables X. We obtain them from forecasting services. There are many forecasting services available, each with its supporters and detractors. We have employed several to use the principle of diversification: from several models we can get a kind of consensus forecast that combines the thinking of all of them while not being overly sensitive to any one. In addition, the spread of the forecasts across the models is a measure of how uncertain is the forecasters' opinion.

We fit five error-correction models for the sales & use tax base, each of the models using different sets of projections from forecasting services. From Global Insight we construct three sets of forecasts based on their baseline, optimistic, and pessimistic projections of national economic indicators. From the Washington State Economic and Revenue Forecast Council (ERFC) we generate forecasts based on their projections of Washington state economic indicators. From the Puget Sound Economic Forecaster (PSEF) we generate forecasts based on their projections of Puget Sound and King County economic indicators.

Variable selection for the models follows a simple methodology. We start with a short list of 5-10 candidate variables. The candidate list is pared down to 1-3 variables using the criteria of forecast mean absolute deviation, the Schwarz criterion, the requirement that $-1 < \gamma < 0$, and that the signs in the cointegrating regression agree with theory. [2] See the Assumptions web page for the variables selected.

Additional Forecasting Models

We use three additional models to bring more stability and information to the distribution of forecasts.

- Regression of (log) King County sales & use tax base on Washington state Sales tax revenue. Forecasts of state tax revenues by ERFC drive the King County tax base forecasts.
- Regression of (log) King County Sales & Use Tax Base on King County taxable sales as compiled by PSEF. Forecasts of King County taxable sales by PSEF drive the tax base forecasts.
- Exponential smoothing of monthly (log) King County Sales & Use Tax Base using the Holt-Winters three-parameter model.[3] A monthly tax base series is imputed from transit tax revenues and then forecasted. Monthly forecasts take into account the seasonality of taxable sales. They are then aggregated into annual forecasts.

Constructing a 65% Confidence Forecast

The eight sets of forecasts exhibit large cross-sectional volatility (see Chart 1), which is a reflection of the uncertainty about where this difficult economy is headed. In particular, the three Global Insight models are in a zigzag pattern that seems to over-correct from year to year, very bearish one year and bullish the next year.

---

[2] Schwarz, G. (1978), "Estimating the Dimension of a Model," *Annals of Statistics,* 6, 461-464
[3] Johnson, Montgomery and Gardiner (1990), *Forecasting and Time Series Analysis,* McGraw-Hill, 2nd Ed.

We make two modifications to the distribution of forecasts. First, we eliminate the pessimistic and optimistic Global Insight forecasts; we also eliminate the PSEF error-correction forecasts. This decision limits the influence of the bearish forecasts for 2010; it is based on knowing that the first four months of sales tax receipts are down only about -2.5% from the first four months of 2009, not the -20% or even less forecasted by these models. Second, we use the PSEF taxable sales forecast model as the mean of the distribution. This is very close to the arithmetic mean of the forecast distribution, but smoother, which is desirable for budget planning.

The mean forecast has a 50% confidence level, meaning there is equal chance of actual revenues received being above or below forecasted revenues. The Forecast Council requires a more conservative forecast, one set at a 65% confidence, meaning there is a 65% probability that actual revenues will exceed forecasted. After eliminating the three forecasts mentioned above, we are left with n = 5 forecasts that form a distribution at each year. The Student's t-distribution with n-1 = 4 degrees of freedom is assumed. The 0.35 percentile of the distribution produces the 65% confidence forecast.

Chart 1